



Emotion Categorization Using Twitter

Rashmi Welekar, Aarti Karandikar, Shubhangi Tirpude

Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, India

ABSTRACT

Recent interactions between regular users of net can be seen using Micro blogging. Several users share their views on various aspects of life regularly. The micro blogging webs resources are wealthy sources of information for opinion mining as well as sentiment analysis. As a result of micro blogging has appeared comparatively recently, there are some analysis works that were dedicated to this subject.

Using sentiments present in twtws posted by users, organizations can get a quick check non public opinion about their business and about their whole organization structure. Sentiment classifiers for twitter datasets are worked upon in recent times and has observed a variety of results. Sentiment classification will benefit corporations by providing knowledge for analyzing client feedback for product or conducting research. Sentiment classifiers ought to be ready to handle tweets in multiple languages[9] to hide a bigger portion of the obtainable tweets.

In this paper we have developed a sentiment classifier which can tell positive, negative and neutral sentiment for a tweet. Experimental evaluations show that our projected techniques are efficient and perform higher than techniques found in literature. We here propose to mine sentiments from micro blogging service, Twitter, where users blog for real time situations. They express their opinions for almost every domain. Throughout this paper, we present a hybrid approach for exploitation lexicon based technique to find out the language point of direction of the opinion words in tweets.

Key words: Micro blogging, Emotion Categorization, Sentiment Analysis, Twitter Analysis

1. INTRODUCTION

Current rising trend of collaborating over the network have increased the chances of communication, sharing resources and collaborative work. Various social networking sites like Twitter, LinkedIn, Snapchat, Instagram, Facebook, and YouTube have gained so popularity that we must take their cognizance. They have established themselves in every arena of the prime applications of internet. With their help users can create their own social network according to their need in a very easy and simple way. With the advent of this the users can share various information types like documents, images, wikis and blogs etc. Social networking sites like Twitter it's evident that it has created a pool of public

opinions which are unique about every possible national as well as international entity like social, sport, religious, political, news, healthcare etc.[15], [16],[8]. Twitter has become a viable option for opinion generation and presentation. But at the same time it throws new and altogether different challenges that we could have never imagined. The method to test these opinions is also incompetent and slow for analysis.

Because there are many users are on the social media posting details of their products and or services, these micro blogging [10] websites are thus a very valuable source of sentiments [2] and opinions. This information and datasets are very much source of social studies. We often tend to use data set of twitter. This data set is nothing but the collections messages posted on twitter. Social media and in particular Twitter contains short messages. The variety of the messages can have personal opinions to public statements.

Various political outfits will also be keen to know if public support their initiatives and program or not. Social welfare groups can open up people's views on latest debates. The information on all these aspects can be gathered from micro blogging services where people usually keep writing everyday about every incident that took place and also about what they like/dislike about socio-political scenario and all aspects of their life [3]. In our research we tried to find out how many times micro blogging has been used for sentiment analysis [4][17], and opinion mining [6].

We have used micro blogging platform and Twitter for the following reasons:

- The platforms of Micro blogging are used by individual to voice their opinions on various subjects. So it can be treated as a valuable pool of opinions of individuals.
- Twitter has tremendous variety of tweets and it grows exponentially every day. The collected text is huge form of big random data.
- The user of Twitter ranges from a daily user to CEO s of organizations, celebrities, managers, directors, politicians, Presidents of nations. So we can have a huge mix of text data from totally different socio-cultural background.

2. RELATED WORK

Sentiment analysis of tweets knowledge is taken into account as a way tougher drawback than than of standard text like review documents. This is often partially because of the short length of tweets, the frequent use of informal

and irregular words, and also the fast evolution of language in Twitter. An oversized quantity of labor has been conducted in Twitter sentiment analysis following the feature-based approaches.

The challenges are for the most part attributed to the dominantly informal tone of the small blogging. Pak and Paroubek[7] used the principle the employment micro blogging and a lot of significantly twitter as a corpus for sentiment analysis[1].

1. Micro blogging platforms are used separate individual to raise their concerns on variety of topics [3], so these platforms become a precious supply of people’s reactions.
2. Twitter gathers a large variety of text messages which contains a huge variety of text posts and it develops into a large pool of information collected on a daily basis.
3. The users of Twitter vary from regular user to CEO’s ,managers, directors of organizations to politicians and even presidents of nations. Therefore, a large variety of socio-cultural corpus is generated.
4. Twitter’s audience is consisting of users from several countries.

Survey through the literature substantiates that the strategies of mechanically expanding upon sentiment at the word level[11] fall under the subsequent 2 categories:

5. Dictionary-based approaches
6. Corpus-based approaches

Further, to alter sentiment analysis, completely different approaches are applied to predict the emotions[12] of words, expressions or documents. In our conceive to mine the sentiment from twitter knowledge we tend to introduce a hybrid approach which mixes the benefits of each lexicon & corpus primarily based strategies knowledge.

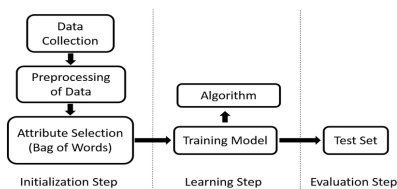


Figure 1: Flow Diagram

Figure 1 shows the flow diagram used to implement our system.

3. METHODOLOGY

The complete operating is split into four steps ,that are

- Take user Input
- Gather knowledge from Internet
- Calculate Sentiment[2]
- Give outline to the user

The raw input can either be provided by the script exploitation some API or it'll be collected from user.once aggregation knowledge from user that's the string (say some-brand) the algorithm/script can attempt to collect the maximum amount as info attainable either by exploitation social network API or by exploitation the

construct of internet scraping. The information is going to be hold on in an array and it'll be passed to the magic rule box which can contain the logic that may calculate the sentiment of every string gift within the array. Once the sentiment has been calculated by the project, it'll offer the outline to the user exploitation some graphs or by exploitation paragraphs.

In order to perform sentiment analysis, we tend to are needed to gather knowledge from the specified supply (here Twitter). This knowledge undergoes numerous steps of pre-processing that makes it a lot of machine[13]wise than its previous type. The following preprocessing mechanisms are applied.

- i. Tweet assortment: Tweet assortment involves gathering relevant tweets concerning the actual space of interest. The tweets are collected exploitation Twitter’s streaming or the other mining tool, for twenty-four hours.
- ii. Pre-processing of tweets: The pre-processing of the information may be a important step because it decides the potency of the opposite steps down in line.
- iii. Removal of re-tweets: The steps concerned ought to aim for creating the information a lot of computer code so as to cut back ambiguity in feature extraction.
- Stop word removal: Stop words that don’t have an effect on the that means of the tweet are removed (for example and, or, still etc.) that checks every word from the text against a lexicon.
- v. Making a dictionary
- vi. Generating a lexicon for words that are necessary
- vii. Twitter feature removal: User names and URLs don't seem to be necessary from the angle of future process, therefore their presence is futile. All usernames and URLs are regenerate to generic tags or removed.
- ix. Stemming: Substitution words with their roots, reducing differing types of words with similar meanings. This helps in reducing the spatial property of the feature set.[5]
- x. Feature Extraction :

A feature may be a piece of knowledge which will be used as a characteristic which might assist in resolution a retardant (like prediction). The standard and amount of options is incredibly necessary as they're necessary for the results generated by the chosen model. Frequency analysis may be a technique to gather options with highest frequencies utilized in. Further, they removed a number of them because of the presence of words with similar sentiment (for example happy, joy, rapturous etc.). The work of extracting options may be a bit completely different as they don’t concentrate on a specific topic or event however proposes to seek out trending topics during a region.

4. APPLICATIONS

- Commerce

Companies can make use of this research for gathering public opinion related to their brand and products. From the company’s perspective the survey of target audience is imperative for making out the ratings of their products. Hence Twitter can serve as a good platform for data collection and analysis to determine customer satisfaction.

- Politics

Majority of tweets on Twitter are associated with politics[14]. Because of Twitter’s widespread use, several politicians are reaching to hook up with individuals through it. Individuals post their support or disagreement towards government policies, actions, elections, debates etc. therefore analyzing knowledge from it will assistance is in determinant public read.

- Sports Events

Sports involve several events, championships, gatherings and a few controversies too. Many folks are ardent sports followers and follow their favourite player’s gift on Twitter. These individuals of times tweet concerning completely different sports connected events. We will use the information to assemble public read of a player’s action, team’s performance, official selections etc.[1]

5.RESULT AND DISCUSSION

Figure 2 shows the applications interface. Fig. 3 shows the top 10 trends from twitter taken on a random day.

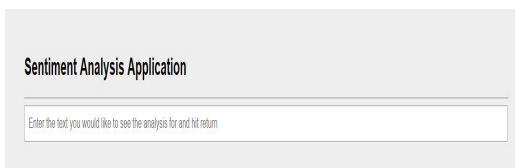


Figure 2: Application GUI

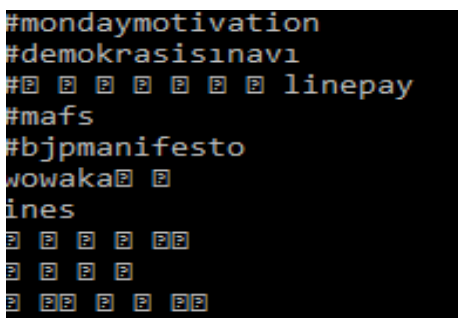


Figure 3: Top Ten Trends



Figure 4: Fetched Tweet

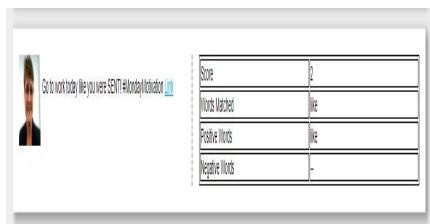


Figure 5: Actual Tweet with Time

Figure 4 shows the tweet fetched by the developed system and Figure 5 shows the actual tweet. Table 1 discusses the results generated on latest trends.

Table- 1: Table of Result

Domain	Trends	Sentiment score	Classification
Politics	#BJP	3	Positive
	#Congress	-1	Negative
Sports	#SachinTendulkar	4	Positive
	#RCB	-2	Negative
Commerce	#OnePlus7	4	Positive
	#Flipkart	2	Positive
Education	#Rcoem	5	Positive
	#ArtificialIntelligence	4	Positive
Bollywood	#Kalank	-2	Negative
	#HowsTheJosh	4	Positive
Social Issues	#meToo	-1	Negative
	#firstVote	3	Positive
Research	#blackHole	2	Positive
	#globalWarming	1	Positive

6. CONCLUSION

Micro blogging these days has become top trend in recent types of the communication. A latest analysis has discovered it as on-line viva-voce disapproval. The large number of information and knowledge present in micro blogging web resources has made a very interesting arena for supply of knowledge for opinion mining and sentiment analysis. In our research work, we have used a corpus [1] which has been trained as a sentiment classifier. Our classifier tool is ready to differentiate between positive, negative and neutral sentiments of documents.

The approach used by us is a completely unique approach for sentiment analysis for knowledge collected from Twitter. The opinion words from the tweets are extracted to find the real sentiment. The technique used for language orientation was corpus based. The technique used for language orientation of verbs and adverbs is dictionary based.

REFERENCES

1. T. Pedersen. **A Simple Approach to Building Ensembles of Naive Bayesian Classifiers for Word Sense Disambiguation.** In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference, USA*, pp. 63–69, May 2002.
2. S. Batra and D. Rao. **Entity Based Sentiment Analysis on Twitter.** Stanford University, 2010

3. R. Parikh and M. Movassate. **Sentiment Analysis of User-Generated Twitter Updates using Various Classification Techniques.** *CS224N Final Report*, April 2009.
4. A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau. **Sentiment Analysis of Twitter Data.** *Proceedings of the Workshop on Languages in Social Media*. pp. 30–38,
5. B. Pang and L. Lee. **Opinion Mining and Sentiment Analysis.** *Foundations and Trends in Information Retrieval*. Pp. 1-135, January 2008
<https://doi.org/10.1561/1500000011>
6. B. Pang, L. Lee and S. Vaithyanathan. **Thumbs up? Sentiment Classification Using Machine Learning Techniques.** *EMNLP*, pp. 79–86, June 2002
<https://doi.org/10.3115/1118693.1118704>
7. A. Pak and P. Paroubek. **Twitter as a Corpus for Sentiment Analysis and Opinion Mining.** *Proceedings of LREC*, pp.1320–1326, January 2010.
8. K. Dave, S. Lawrence and D. Pennock. **Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews.** *Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews*, pp. 519–8, October 2003.
<https://doi.org/10.1145/775152.775226>
9. H. Schmid. **Probabilistic Part-of-Speech Tagging Using Decision Trees.** *Proceedings of International Conference on New Methods in Language Processing*. April 2006
10. C. Shannon. **A Mathematical Theory of Communication.** *Mobile Computing and Communications Review*. pp. 3-55, January 2001
<https://doi.org/10.1145/584091.584093>
11. T. Wilson, J. Wiebe and P. Hoffmannl. **Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis.** *Proceedings of HLT/EMNLP*, October 2005.
12. C. Yang, K. Lin and H. Chen. (2007). **Emotion Classification Using Web Blog Corpora.** *International Conference on Web Intelligence, Washington*, pp.275-278, September 2007
<https://doi.org/10.1109/WI.2007.51>
13. E. Alpaydin. **Introduction to Machine Learning (Adaptive Computation and Machine Learning).** *The MIT Press*, October 2004
14. H. Anthony. **Probability and Statistics for Engineers and Scientists.** *Duxbury, Belmont, CA, USA*, 2007
15. S. Tirpude, **Abnormal X-Ray Detection System using Convolution Neural Network.** *International Journal of Advanced Trends in Computer Science and Engineering*, pp.828-832. May 2020
<https://doi.org/10.30534/ijatcse/2020/119912020>
16. A. Karandikar ,V. Deshpande,S. Singh, S. Nagbhidkar and S. Agrawal . **Deep Flake Video Detection using Neural Network.** *International Journal of Advanced Trends in Computer Science and Engineering*, May 2020
17. J. Read. **Using emoticons to reduce dependency in machine learning techniques for sentiment classification.** *In The Association for Computer Linguistics*, 2005
<https://doi.org/10.3115/1628960.1628969>